



吕勇稷

西湖大学工学院

✉ lyuonjji@westlake.edu.cn

📍 浙江省杭州市

📞 13213725153

🌐 <https://memory-source.github.io/>

专业技能

- **编程语言**: Python (熟练), NumPy、Pandas、Matplotlib、PyTorch、Scikit-learn。
- **机器学习与深度学习**: 深入理解 CNN、RNN、Transformer、Attention 机制等经典架构; 扩散模型 (DDPM、D3PM、SEDD)、对比学习 (CLIP)、图神经网络 (GAT、GCN)。
- **跨学科应用**: 熟悉机器学习在生物学 (基因组学、蛋白质结构预测) 中的应用。
- **开发工具**: Git、Linux 命令行、Docker、PyTorch。
- **AI 编程工具**: Claude Code 等 AI 辅助编程工具。
- **可视化工具**: ComfyUI、Matplotlib。
- **科研工具**: LaTeX、Jupyter Notebook、ArXiv、Papers with Code。
- **信息追踪**: 自建工作流, 每日追踪并阅读 arXiv 等平台的最新 AI 研究论文并推送到自己飞书账号, 保持对前沿技术的敏锐度。
- **英语能力**: 雅思备考中; 可流畅阅读英文论文和技术文档; 西湖大学全英授课。

工作经验

- | | | |
|---|----|--------------|
| 西湖大学AGI Lab | 成员 | 2025/08 - 至今 |
| 前沿技术研读与复现: 系统研读扩散语言模型 (D3PM、SEDD)、多模态图神经网络、对比学习 (CLIP)、DDPM 等方向前沿论文, 完成从数学推导到代码复现的完整学习路径。 | | |
| • 多模态组学数据整合方法调研: 参与组学数据 (RNA/蛋白/Peak 多模态) 与图注意力网络结合的方法调研, 梳理联邦学习与隐私计算在 GWAS 场景下的应用文献。 | | |
| 西湖大学 Pebble 暑期学校 | 助教 | 2026/07 - 至今 |
| 主讲“机器学习在生物学中的应用”课程, 面向药学及生物学博士生介绍 AI 在基因组学、蛋白质结构预测等领域的交叉应用。 | | |
| • 设计互动式教学内容, 制作可视化讲义, 帮助学生理解复杂的机器学习概念 | | |

项目经历

- | | | |
|---|-------|-------------------|
| STAMP: Spatial Multi-Omics Fusion Algorithm | 独立开发者 | 2026/04 - 2026/07 |
| https://github.com/memory-source/Stamp-A-algorithm-about-Spatial-Multi-omics | | |
| 项目简介: 基于图神经网络的空间多组学融合与聚类方法, 解决空间转录组学与 ATAC-seq 数据整合中的异质性问题。 | | |
| 核心创新: | | |
| • 双图 GAT 编码器: 设计共享权重的图注意力网络, 同时处理空间图 (kNN on 坐标) 和特征图 (kNN on PCA/LSI), 通过可学习权重融合。 | | |
| • 非对称跨模态注意力: 仅 RNA 查询 ATAC (单向性), 保护高质量 ATAC 模态免受噪声 RNA 污染。 | | |
| • 跨模态重建: 每个潜在表示必须重建两个模态 (4 个 MSE 项), 防止表示崩溃。 | | |
| 技术成果: | | |

- 在 5 个模拟数据集上平均 ARI 达到 0.903，在 10 种方法中排名第 3。
- 在真实小鼠脑数据（P22）上验证，ARI = 0.504 vs cortical-layer annotations。
- 核心实现约 260 行代码，训练稳定（400–800 epochs 早停）。

技术栈：Python、PyTorch、图神经网络（GAT）、多模态学习、空间转录组学、ATAC-seq、聚类分析

Knowpath Agent Orchestration

核心开发者

2026/3 - 2026/07

<https://github.com/lpz2020/knowpath>

基于 FastAPI + Pydantic 开发约 11,000 行 Python 后端，实现统一的 BaseAgent 基类与 perceive → plan → act 三阶段执行模板，通过 AgentTrace 记录动作序列、LLM 调用次数与毫秒级延迟，实现全链路可审计。

- 实现 StudentDiagnosisAgent：设计 EMA 多源证据融合状态机，将测验 / 对话 / 自我解释三类信号按可信度增益（ $\alpha = 0.30/0.12/0.20$ ）聚合为 Bloom 六级掌握度状态，并落地“LLM 只产出证据分数、状态转移由确定性代码执行”的分工，规避 LLM 直接裁决学情的不可靠性。
- 实现 LearningPathAgent：先修 DAG 硬约束（反向 BFS 收集先修闭包 + 掌握度过滤 + Kahn 拓扑排序）× 五因子软排序，冷启动下全程 0 次 LLM 调用，每步排序可溯源至具体因子得分。
- 实现 PersonalizedTutorAgent：以先修缺失数（0/1–2/>2）作为纯函数切换 Bridge / Scaffold / Remediate 教学策略，将“知识诅咒”落地为可计算机制。
- 系统在 7 门真实课程上自动构建 571 个概念、156 条关系，24 项模块级测试全部通过。

技术栈：Python、FastAPI、Pydantic、LLM Agent 架构、知识图谱、DAG 拓扑排序、Bloom 分类法、EMA 状态机

教育经历

西湖大学

人工智能 · 本科在读

2025/08 - 2029/06

相关课程：线性代数、微积分、程序设计（C/Python）、机器学习基础、深度学习、数据结构与算法。

自主学习：

- 系统完成《动手学深度学习》（李沐著）课程学习，深入理解 CNN、RNN、Transformer 等核心架构，完成全部代码实战，并按教材进度参加 Kaggle 房价预测实战。
- 完成李宏毅《机器学习》系列课程，对模型训练、优化算法、注意力机制等有深刻理解。